A Digital Duty of Care for Australia—Developing a duty of care framework for online services used by Australians

Preamble

The government is developing draft legislation to implement a Digital Duty of Care in Australia under the *Online Safety Act 2021* (the Act). This document sets out the approach the Government is taking to establish the duty of care framework that will apply to providers of online services to Australians.

This paper reflects feedback from previous discussions held with stakeholders and decisions taken by Government in developing its response to the independent statutory review of the Act, which was released on 14 April 2026. In developing a Digital Duty of Care framework, the Government has been guided by the following principles that would underpin the duty:

1. improving and promoting online safety for Australians
2. providing clarity for regulated entities and minimising regulatory complexity
3. building upon the current online safety protections for Australians
4. improving accountability and transparency
5. building in sufficient flexibility to accommodate technological change.

The duty of care reforms will complement the Social Media Minimum Age law and reforms underway to modernise the National Classification Scheme.

About the Digital Duty of Care

The Digital Duty of Care (the Duty) will support broad, risk-based and proportionate regulation of all online service providers.

The details included in this paper reflect current considerations and approaches. They are outlined as the basis for further consultation and refinement.

The Duty will require all entities regulated by the Act to maintain effective systems and processes that:

- so far as is reasonably practicable, provide a safe online environment for all Australians
- prevent, monitor and appropriately address content and activity that is illegal or harmful to young people, and
- Ensure the safety of service features, including AI and algorithmic content recommendation or generation systems and bot accounts.

The failure of a service provider to take reasonable steps to maintain effective systems and processes will constitute a breach of the Duty.

At its core, the Duty recognises that online service providers are best placed to manage risks of serious harm associated with use of their service. The Government will hold online service providers to account by requiring that they undertake thorough and regular risk assessments, implement effective mitigation strategies and measures, and report transparently on the effectiveness of those measures in driving safer online experiences for all Australians.

The Duty of Care will place a proactive obligation on industry

To uplift safety online, the Duty will place the responsibility for protecting the online safety of Australians on online service providers, recognising that they are best placed to assess and mitigate the risk of serious harms on their own services. The Duty will achieve this through:

- Comprehensively updating the proactive elements of the Act to embed safety by design and require systemic and preventative action by services 'upstream' of the kind of seriously harmful behaviour and content captured by the Act's existing complaints-based removal schemes, which will be retained.
- Requiring ongoing due diligence from services, to assess and mitigate risk regularly and as their services – and the potential harms that could eventuate from their services – change.

The Duty of Care will be focused on systems and processes

The Duty will focus on ensuring that services have effective systems and processes in place that are reasonable, risk-based and proportionate, and which mitigate the risk of reasonably foreseeable serious harm to their users. Services will not be liable for individual instances of harmful content or activity but will be required to act reasonably and in good conscience to prevent reasonably foreseeable harms from occurring. However, frequent instances of harmful content or activity on a service, or a service's failure to appropriately address instances of harmful content or activity, may be considered evidence of breach of the Duty.

Outside of the existing complaint schemes, eSafety will not investigate individual pieces of content or seek to arbitrate cases on behalf of individuals except where it is necessary to investigate, assess or conduct compliance activities in relation to systemic failures of the Duty of Care.

The Duty will apply to all online services currently within scope of the Online Safety Act

The Duty will be broad based, applying to all online service providers currently defined as within the scope of the Act. This means the Duty will capture services throughout the 'tech stack', including:

- Social media services (including age-restricted social media platforms)

- Relevant electronic services, such as messaging apps, interactive online games, and online dating services
- Designated internet services, such as general websites or apps, many online pornography services, and many generative AI services (including chatbots and image generators)
- Hosting services, which host data and material provided by other online services
- Internet service providers which connect users' devices to online services
- Search engine services, which index and recommend online services and resources in response to user queries
- App distribution services, or 'app stores' which index and recommend apps for download onto user devices
- Equipment and operating system services, such as phones, tablets, laptops, wearable devices, and the operating systems that make them function.

The obligations of the duty will be proportionate to a service's risk profile and their opportunity to address risks in the chain of service provision to users. Each service would be required to assess their risks and determine what preventative responses, systems and processes they can and ought to develop and implement to meet the Duty.

The Duty will be underpinned by Government made rules subject to Parliamentary scrutiny

Under the Duty, the Minister for Communications (the Minister) or, at the Minister's direction, the eSafety Commissioner (eSafety), will be given the power to make binding rules for:

- designating additional categories and types of harm, as being within the scope of the Duty's obligations
- specifying compliance requirements for service providers, or specified classes of services, with the Duty
- specifying services, or classes of services, that are exempt from the Duty's obligations.

Any rules made by the Minister, or eSafety (at the Minister's direction), will be subject to parliamentary scrutiny as a disallowable legislative instrument and will be required to include a statement of the rule's compatibility with Australia's human rights obligations, including freedom of expression.

The Duty will complement, not replace, the Act's complaint schemes

The Act's existing complaints schemes including cyberbullying, adult cyber abuse, image-based abuse and illegal or restricted material will remain in place.

Amendments will be made to the Act to establish a link between services' compliance with other statutory requirements in the Act, including notices under the complaints scheme, and their compliance with the Duty.

Regard to the rights and freedoms of Australians

In meeting their obligations under the Duty, service providers will be expected to have regard to Australians' rights and freedoms, including freedom of expression and the rights of children.

As outlined above, any binding rules made by the Minister, or eSafety (at the Minister's direction), under the Duty will be required to demonstrate their compatibility with human rights and be subject to scrutiny by Parliament as disallowable legislative instruments.

The Digital Duty of Care

Part 1: Principal elements of the Duty

As noted above, the basis for consultation is that the Digital Duty of Care will require service providers to maintain **effective systems and processes** that:

- so far as is reasonably practicable, provide a **safe online environment** for all Australians
- prevent, monitor and appropriately address content and activity that is **illegal or harmful to young people**, and
- ensure the safety of **service features**, including **AI and algorithmic content recommendation or generation systems and bot accounts**.

The failure of a service provider to take reasonable steps to maintain effective systems and processes will constitute a breach of the Duty.

Systems and processes

To meet the Duty, service providers will be obliged to maintain a range of effective systems and processes. This means that service providers will have to put in place effective risk management systems, assess services to identify potential risks of harm, and to put in place strategies or measures to prevent or mitigate reasonably foreseeable serious harms. This process will require regular assessment and review, including to capture changes to services and their product offerings.

Service providers will be required to evaluate the effectiveness of the prevention and mitigation strategies and measures put in place. Effective systems and processes will necessarily vary between services and it will be up to services to determine what is appropriate and sufficient. However, examples of systems and processes could include:

- content classification or moderation systems, tools and processes for preventing, detecting, reporting, and/or appropriately addressing potentially harmful material and activity
- systems and processes for engaging with law enforcement and public authorities concerning matters such as illegal activity (including threats to public safety)
- systems and procedures for internal dispute resolution
- terms of use, policies, community standards or standards of conduct that adequately cover harmful material and activity on the service, and are effectively enforced
- user empowerment tools and controls, and user account settings, that allow them to tailor privacy, safety and recommendation settings
- parental controls and family safety settings.

A safer online environment

The following three overarching principles of safety by design will guide the development of the regulatory framework:

- **Service provider responsibility**: service providers take active responsibility for the safety and safe use of their services, understanding, assessing and addressing online harms in the design and provision of those services
- **User empowerment and autonomy**: service providers centre the best interests, dignity and rights of their users in the design and operation of their services
- **Transparency and accountability**: service providers are transparent in their approach to safety in the design and operation of their services, and accountable for this approach – including ongoing innovation and sharing of good practice in online safety.

Safety by design is a well-established concept, and eSafety has proven expertise and experience in providing research and guidance on how industry can embed safety by design principles in the design, development, deployment and operation of their services.

Illegal and harmful content and activity

Content and activity that is the subject of the Duty includes that which is illegal or harmful to young people. The Minister for Communications would have the ability to determine, by disallowable legislative instrument, additional harms as being within the scope of the Duty.

Illegal content and activity **includes** content and activity that:

- is Class 1 material under the Act, or would be Refused Classification under the National Classification Scheme, including child sexual exploitation material (CSEM), pro-terror material, and material that promotes, incites or instructs in matters of crime or violence
- amounts to a criminal offence under Commonwealth law including causing or threatening psychological and physical harm, menacing and harassing communications, doxxing and stalking, image-based abuse and unsolicited sending of sexual material (e.g. cyber-flashing), and hate crime offences.
- presents a seriously harmful threat to Public Safety at an appropriate threshold that preserves individual's freedom of expression and free speech.

Content and activity that is harmful to young people and **includes**:

- content and activity that promotes or provides, for example, instruction in eating disorders, which may be especially harmful to young people
- content and activity that promotes or glorifies seriously harmful activity by young people such as dangerous stunts or promotion of crimes
- Class 2 material under the Act including online pornography and other material that is not age appropriate for children, such as self harm and suicide material.

Features which may increase risk of harm

The Duty will require that service providers proactively assess the potential for reasonably foreseeable serious harms to emerge from the use of service features and to put in place mitigation strategies or measures to prevent or reasonably minimise any serious harm risks. A 'serious harm' threshold will strike a balance between protecting individuals from harm and not unduly impinging their freedom of expression and speech.

The application of safety by design principles to a services' features could include:

- Services that allow the posting of user generated content putting in place measures that prevent the upload of or detect and remove illegal or seriously harmful content such as image-based abuse, CSEM or promotion of violent extremism
- Services providing artificial intelligence tools putting in place measures to prevent the generation and sharing of content by users of illegal or seriously harmful material such as, for example, images of persons 'nudified' without their consent
- The design of algorithmic recommender systems so that they prevent the dissemination of illegal or seriously harmful content or reduce the risk of exposure to seriously harmful behaviours, for example, eating disorders in young people.

Other features that may need to be addressed to meet the requirements of the Duty include those that could **in certain circumstances** enable the reasonably foreseeable creation and interaction of online accounts or identities that cause serious harm. This could include the design of:

- systems that allow automated or 'bot' accounts, which could be used to support or propagate seriously harmful inauthentic activity
- systems that allow anonymous or pseudonymous accounts which could, for example, be misused to support or propagate seriously harmful inauthentic activity or to menace, threaten and harass other users

- features that allow discovery of or connection with other accounts, which may enable illegal or seriously harmful activity such as, for example, grooming.

Reasonable steps would require services to have regard to the legitimate uses and functions of these features and peoples' right to freedom of expression so that these are not unduly impaired or undermined.

Part 2: Core obligations under the Duty

Risk Assessment and Mitigation

Central to the Duty will be an obligation to conduct risk assessments and implement strategies or measures to mitigate reasonably foreseeable identified risks of serious harm. This obligation will require that services:

periodically (at least annually) assess the risk that their service may cause serious harm to users. This would include whenever a significant change is made to the service, such as the introduction, or modification, of service features and changes to terms of service

- implement appropriate and effective mitigation strategies and measures to effectively mitigate potential risks, including safety by design
- undertake regular review and evaluation of the effectiveness of mitigation strategies and measures, and make appropriate changes to improve their efficacy.

Service providers will also be required to consider the best interests of the child in assessing and mitigating risks arising from the design and operation of their services.

Part 3: Compliance and Enforcement

As the regulator, eSafety would have a range of graduated enforcement options to support compliance with the Duty. It is intended that services will generally be given the opportunity, in the first instance, to identify and rectify deficiencies in their systems and processes.

In terms of escalation within the graduated framework, eSafety's actions will be informed by the severity of the non-compliance and this will be supported by a range of formal powers given to eSafety, including:

- information gathering and reporting notice powers
- formal warnings
- remedial directions and enforceable undertakings
- audit powers
- infringement notices
- application to the Courts for civil penalty orders or injunctions.

Reporting obligations to eSafety and the public

Under the Duty, existing reporting notice powers in the Act will be adapted. At a high level, services will be required to provide information to support eSafety's regulatory oversight function. This may include:

- periodic reporting obligations for specified services or classes of services (selected according to objective and risk-based criteria), requiring regular reporting to eSafety by such services on how they are complying with requirements of the Act, and reporting on information that may be specified in the rules
- more targeted one-off (non-periodic) reporting notices by services or classes of services prescribed by eSafety
- making other information publicly available as required by rules.

There will be appropriate safeguards in place concerning the collection and handling of confidential and commercial-in-confidence information for the purposes of assessing whether a service has met its Duty

obligations. Services will not be permitted to withhold information from eSafety on the basis of commercial in confidence status.

Audit obligations

eSafety will have the power to direct a service to procure and provide an independent audit, at the service's own expense, of the effectiveness of their systems and processes to comply with the Duty (e.g. of their risk management practices). Such audits could only occur where eSafety had a reasonable and defensible basis to believe that a service was failing to systemically meet its duty obligations.

Protection for researchers

There is an important role for independent researchers and research institutions in assessing or auditing the safety practices and user experiences on services. Consideration is being given on how to protect the role of authorised independent researchers. Protections would allow greater transparency over the practices of online service providers.

Penalties

Enforceable provisions will be graduated but include serious civil penalties for egregious and systemic breaches of the Duty. In terms of penalty units, the maximum civil penalty will be commensurate to offences under other legislation such as the *Competition and Consumer Act,* where maximum penalties for ACCC enforcement were recently increased to $100 million.

High civil penalties are intended to serve as a deterrent and would only be applied in circumstances where there has been egregious or systemic non-compliance. Where non-compliance has occurred, the eSafety Commissioner will be empowered to employ other remedial measures as part of graduated powers to promote service compliance with the Duty.

Part 4: Transition

The Duty will incorporate and adapt existing elements of the Act, initially the BOSE and, potentially, over time, the industry codes or standards made under the Online Content Scheme. This will necessitate appropriate transitional arrangements, which will be worked through as part of developing draft legislation.

The approach to transitioning to a new regulatory framework will be grounded in these main principles:

- Reducing complexity and regulatory burden wherever possible
- Creating flexibility, and wherever possible making sure rules can be applied so as to best fit the diverse circumstances of regulated services
- Ensuring the current safety standards are not reduced, with the existing codes and standards serving as a baseline.

A period of 12 months is proposed between the legislation passing and commencement of the Duty. During that period, the necessary guidance to support the implementation of the Duty would be developed and made available. The 12-month transition period will allow sufficient time for this guidance to be issued and, in turn, ensure services are prepared for commencement of the Duty.

Other Online Safety Act Amendments

The following recommendations from the statutory review of the Online Safety Act will also be addressed in the Bill. In the Government's response to the Review, these recommendations have been supported or supported in principle. Where recommendations are supported in principle, their implementation through the Omnibus Bill may differ from the recommendations but will deliver on the intended outcome.

Recommendation number	Recommendation
Recommendation 1 (Support)	That the objects of the Act should be amended to include more descriptive objectives that are linked to the various functions covered by the Act.
Recommendation 4 (Support)	That Australia adopt a singular and overarching duty of care that encompasses due diligence, and is underpinned by safety by design principles, risk assessment, risk mitigation and measurement.
Recommendation 5 (Support in principle)	The harms that should be highlighted for attention under a duty of care should at a minimum include: • Harms to young people, including child sexual exploitation and abuse (including grooming), bullying and problematic internet use • Harms to mental and physical wellbeing, including threats to harm or kill, or attacks based on a person or group of people's protected characteristics, such as sex, gender, sexual orientation, race, ethnicity, disability, age or religion • Instruction or promotion of harmful practices, such as self-harm/suicide, disordered eating and dares that could lead to grievous harm • Threats to national security and social cohesion, such as through promotion of terrorism and abhorrent violent extremist content; and • Other illegal content, conduct and activity.
Recommendation 6 (Support in principle)	**Entities with the greatest reach or risk should be required to complete a risk** assessment at least every 12 months and to carry out a risk assessment when significant changes are made to the design and operation of their service. These entities should also be required to provide an annual report detailing their risk assessments, risk mitigations and how successful they have been to the regulator.
Recommendation 8 (Support)	The best interests of the child should be a primary consideration for online service providers in assessing and mitigating the risks arising from the design and operation of their services, including risks to children who may use the service and risks to children as a result of how the service may be used.
Recommendation 9 (Support in principle)	The eSafety Commissioner should be empowered to create mandatory rules (in the form of codes) on how entities can comply with certain aspects of the duty of care requirements, including addressing specific online harms. This should not stop services from taking additional steps to protect people. Codes would not create safe harbours.
Recommendation 10 (Support in principle)	In addition to risk assessments, a service with the greatest reach or risk should be required to provide an annual transparency report and publish a summarised version on its website. This should not replace the broad power for eSafety to require periodic and non-periodic transparency reports from all services.
Recommendation 12 (Support)	The regulator should have the discretion and power to require services to undertake an audit at their own expense.

Recommendation number	Recommendation
Recommendation 13 (Support in principle)	Subject to adequate safeguards, services with the greatest reach or risk should be required to share data with authorised researchers for the purposes of determining compliance with a duty of care model, the takedown schemes and research into emerging problems and harms.
Recommendation 15 (Support)	Users experiencing adult cyber abuse or child cyberbullying should only need to wait 24 hours (not 48 hours) following a complaint to a service before eSafety is able to issue a removal notice.
Recommendation 16 (Support)	The regulator should be empowered to waive the statutory delay to issue a removal notice for the child cyberbullying and adult cyber abuse schemes where no clear complaint mechanism exists on the online service, or where reporting would lead to a reasonably foreseeable risk of further harm to the user experiencing the abuse.
Recommendation 19 (Support)	The Act should enable the regulator to issue a removal notice for material that has met the regulatory threshold for removal under a prior complaint, where the regulator becomes aware that the material has been reposted.
Recommendation 25 (Support)	All services should be required to have an easily accessible, simple and user-friendly way to make a complaint and internal complaint handling processes that are in line with a code on internal dispute resolution. In particular, this should include a way for non-users to report issues such as when intimate images have been posted without consent on a service. Services should also be required to respond to reports within a reasonable time and for some issues within 24 hours.
Recommendation 27 (Support)	The government should explore how best to prohibit search engines and app stores from surfacing, selling or distributing 'nudify' apps and undetectable stalking apps.
Recommendation 34 (Support)	The maximum civil penalty that a court can impose should be increased to the greater of 5 per cent of global annual turnover or $50 million.
Recommendation 35 (Support)	The civil penalties for non-compliance with removal notices should be increased to a maximum of $10 million for companies.
Recommendation 36 (Support in principle)	The Act should be amended to empower the regulator to use enforceable undertakings or issue remedial directions to services in relation to all relevant penalty provisions, to seek to bring them back into compliance.
Recommendation 37 (Support)	The Act should allow removal and link-deletion notices to be issued simultaneously under the Online Content Scheme.
Recommendation 38 (Support)	The Act should empower the regulator to simultaneously issue link removal notices for all harmful content under removal schemes.
Recommendation 39 (Support)	The finalised duty of care model should include scope to consider repeated non-compliance by services in removing content as evidence of non-compliance with the duty of care.

Recommendation number	Recommendation
Recommendation 44 (Support)	The Act should require major platforms, that is those designated under the reach or risk criteria under the duty of care requirements, to have a contact point for service in Australia.
Recommendation 46 (Support in principle)	The Act should be amended to empower the regulator with stronger powers in relation to investigations, including to: • Incorporate the monitoring and investigations provisions of the Regulatory Powers Act into the Act • Initiate investigations of a service's compliance with the duty of care; and Initiate investigations into reposted material that was previously reported and taken down.
Recommendation 47 (Support in principle)	Amend the Act to provide the regulator with appropriate flexibility to conduct investigations as it thinks fit, including the use of technological tools to assist with investigations and content removal, and the use of sock-puppet accounts.
Recommendation 50 (Support)	Section 205 of the Act should be amended to confirm that non-compliance with a requirement to give evidence includes information as requested under section 199 (and other sections in Part 14 of the Act).
Recommendation 52 (Support)	The Act should be amended to require services to maintain certain records, such as measures taken to comply with obligations under the Act and any actions taken in response to the regulator's requests and risk assessments, for the purposes of eSafety's investigations.
Recommendation 66 (Support)	That the updated Act be subject to independent review within three years of the commencement of the key reforms to the Act, or by 2029, whichever is earliest.